UFRJ - CCMN - IM - Departamento de Métodos Estatísticos
Probabilidade e Estatística - Estatística

Prova # 02 09-07-2013

Atenção: Não serão aceitas respostas sem justificativa: as expressões que levaram a alguma resposta numérica devem ser indicadas nos espaços apropriados.

1. Uma empresa deseja estudar a influência da propaganda sobre a venda de seus produtos. Para isso, obteve os seguintes dados relativos a 5 produtos seus:

X	10	20	30	40	50
Y	10	10	20	20	40

onde X é o investimento em propaganda (em milhares de reais) e Y é a o valor da venda do produto (em milhões de reais). Pede-se:

(a) Correlação entre X e Y; um gráfico de dispersão de X e Y;

(b) Valores estimados (por mínimos quadrados) dos parâmetros da reta $Y = a + bX$;

(c) Valor estimado da venda quando a propaganda é 35 mil reais. Forneça uma interpretação do coeficiente b no contexto deste problema.

Solução:

(a) $\sum x_i = 150$, $\sum x_i^2 = 5500$, $\sum y_i = 100$, $\sum y_i^2 = 2600$,
$\sum x_i y_i = 3700$,
$cor(X,Y) = \frac{3700 - \frac{150\times 100}{5}}{\sqrt{(5500-\frac{150^2}{5})\times(2600-\frac{100^2}{5})}} = 0.9037$;

(b) $b = \frac{3700-\frac{150\times100}{5}}{5500-\frac{150^2}{5}} = 0,7$; $a = \bar{y} - 0,7 \times \bar{x} = \frac{100}{5} - 0,7 \times \frac{150}{5} = -1$;

(c) se $x_0 = 35$ mil reais, $y_0 = -1 + 0,7 * 35 = 23,5$ milhões de reais.
O coeficiente b mede o quociente $\frac{\Delta y}{\Delta x}$, isto é, a variação da receita da venda sobre a variação do investimento em propaganda. Ou seja, neste contexto, em média para cada 1000 reais a mais investido em propaganda haverá um aumento de 700.000 reais nas vendas.

2. Definimos a variável $E = \bar{X} - \mu$ como sendo o erro amostral da média, onde $\bar{X}$ é a média de uma amostra aleatória de tamanho n de uma população com distribuição Normal de média μ e desvio padrão $\sigma = 20$.

(a) Que proporção das amostras de tamanho 100 terá o valor absoluto do erro amostral maior do que 2 unidades?

(b) Neste caso, qual deve ser o valor de δ para que $P(|E| > \delta) = 0,01$?

(c) Qual deve ser o tamanho da amostra para que o módulo do erro amostral seja inferior a 1 unidade com 95% de probabilidade?

Solução:

(a) $P(|E| > 2) = 1 - P(|E| \leq 2) = 1 - P\left(\frac{-2}{20/10} \leq Z \leq \frac{2}{20/10}\right) = 2(1-\phi(1)) = 0,3174.$

(b) $P(|E| > \delta) = 0,01 \Rightarrow P\left(|Z| \leq \frac{\delta}{2}\right) = 0,99 \Rightarrow \Phi(\frac{\delta}{2}) = 0,995 \Rightarrow \delta = 2 \times 2,58 = 5,16.$

(c) $n = \left(z_{1-\frac{\alpha}{2}}\sigma/d\right)^2 = \left(\frac{1,96\times20}{1}\right)^2 = 1536,64; assim, n = 1537.$

3. Os candidatos A e B disputam as eleições para a prefeitura da cidade de Bayesville. Em uma pesquisa de intenção de voto, uma amostra aleatória de tamanho 400 foi retirada da população e observou-se que a proporção dos eleitores que votariam em A era de $0,55$ e a proporção dos eleitores que votariam em B era de $0,30$, sendo o restante brancos, nulos e indecisos.

   (a) Determine intervalos de 90% de confiança conservativos para a proporção de eleitores na população que votariam em A e para a proporção de eleitores na população que votariam em B, cada um deles com 3 casas decimais;

   (b) Usando novamente a opção conservativa, qual deve ser o tamanho da amostra para que a margem de erro para cima ou para baixo da proporção amostral dos eleitores que votariam em A e da proporção amostral dos eleitores que votariam em B seja de 0,02 (ou de 2 pontos percentuais, usando a linguagem dos noticiários jornalísticos) com 95% de confiança?

Solução:

(a) Candidato A:

$$\begin{aligned} IC(p) &= \left(0,55 - \overbrace{1,64}^{z_{0,95}} \frac{1}{2\sqrt{400}};\ 0,55 + 1,64\frac{1}{2\sqrt{400}}\right) \\ &= (0,509;\ 0,591). \end{aligned}$$

Candidato B:

$$\begin{aligned} IC(p) &= \left(0,30 - 1,64\frac{1}{2\sqrt{400}};\ 0,30 + 1,64\frac{1}{2\sqrt{400}}\right) \\ &= (0,259;\ 0,341). \end{aligned}$$

(b) $z_{1-\frac{\alpha}{2}} \frac{1}{2\sqrt{n}} = 0,02 \Rightarrow \overbrace{1,96}^{z_{0,975}} \frac{1}{2\sqrt{n}} = 0,02 \Rightarrow n = \left(\frac{1,96}{2\times 0,02}\right)^2 \Rightarrow n = 49^2 = 2401.$

4. Um pesquisador está estudando a resistência de um certo material sob determinadas condições. Ele sabe que essa variável é Normalmente distribuída com variância igual a 4 unidades^2. Foi extraída uma amostra aleatória de tamanho 10 obtendo-se os seguintes valores: 7,9 6,8 5,4 7,5 7,9 6,4 8,0 6,3 4,4 5,9. Sendo: $\sum x_i = 66,5;$ *e* $\sum x_i^2 = 455,29$

   (a) Elabore um procedimento para testar a afirmação do fabricante de que a média populacional da resistência é de pelo menos 7,8 contra a alternativa de que ela é estritamente menor que esse valor. Qual a sua conclusão ao nível $\alpha = 1\%$?

   (b) Qual seria o p-valor deste teste?

   (c) Como você procederia, e qual seria a sua conclusão, se no item (a) não fosse conhecida a variância populacional?

Solução: Queremos testar $H_0 : \mu \geq 7,8$ contra $H_1 : \mu < 7,8$.

   (a) Como σ é conhecido, o critério é: Rejeitar H_0 se $Z = \frac{\bar{x}-\mu_0}{\frac{\sigma_0}{\sqrt{n}}} < z_{tab} = z_{0,01}$. Caso contrário, aceitar H_0. Sob H_0, $Z \sim N(0;1)$. Assim, $z_{obs} = \frac{\bar{x}-\mu_0}{\frac{\sigma_0}{\sqrt{n}}} = \frac{6,65-7,8}{2}\sqrt{10} = -1,82$ e
   $P(Z < z_{tab}) = 0,01 = \Phi(z_{tab}) \Rightarrow z_{tab} = -2,33.$
   Então, como $z_{tab} < z_{obs}$, não há evidência para a rejeição da hipótese H_0;

   (b) p-valor$= P(Z < -1,82) = P(Z > 1,82) = 1 - \Phi(1,82) = 1 - 0,9656 = 0,0344;$

   (c) Como σ é desconhecido, o critério é: Rejeitar H_0 se $T = \frac{\bar{x}-\mu_0}{\frac{S}{\sqrt{n}}} < t_{tab} = t_{0,01;9}$. Caso contrário, aceitar H_0. Sob H_0, $T \sim$ t-student com 9 g.l. Então, $P(T < t_{tab}) = 0,01 \Rightarrow t_{tab} = -t_{0.99} = -2,821$
   $s^2 = \frac{1}{n-1}\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right) = \frac{1}{9}\left(455,29 - \frac{(66,5)^2}{10}\right) = 1,4517 = (1,205)^2$
   $t_{obs} = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}} = \frac{6,65-7,8}{\frac{1,205}{\sqrt{10}}} = -3,02.$
   Assim, $t_{obs} = -3,02 < t_{tab} = -2,821$. O que implica em rejeitar a hipótese H_0.

Questão 1-a: Gráfico de dispersão de X e Y